Citation for published version:
Ball, A 2013, 'Introducing the Community Capability Model Framework and White Paper', Community Capability Model Framework for Data-Intensive Research – Applying the Model, Amsterdam, Netherlands, 14/01/13.

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Publisher Rights
CC BY-SA

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

Introducing the Community Capability Model Framework and White Paper

Alex Ball

DCC/UKOLN, University of Bath, UK

14 January 2013

Community Capability Model Framework for Data-Intensive Research: Applying the Model

Workshop 1, IDCC 2013, Amsterdam

Microsoft Research
UNIVERSITY OF BATH
UKOLN
DCC
CC BY SA

Outline

Background

- The Fourth Paradigm
- Project objectives

The White Paper

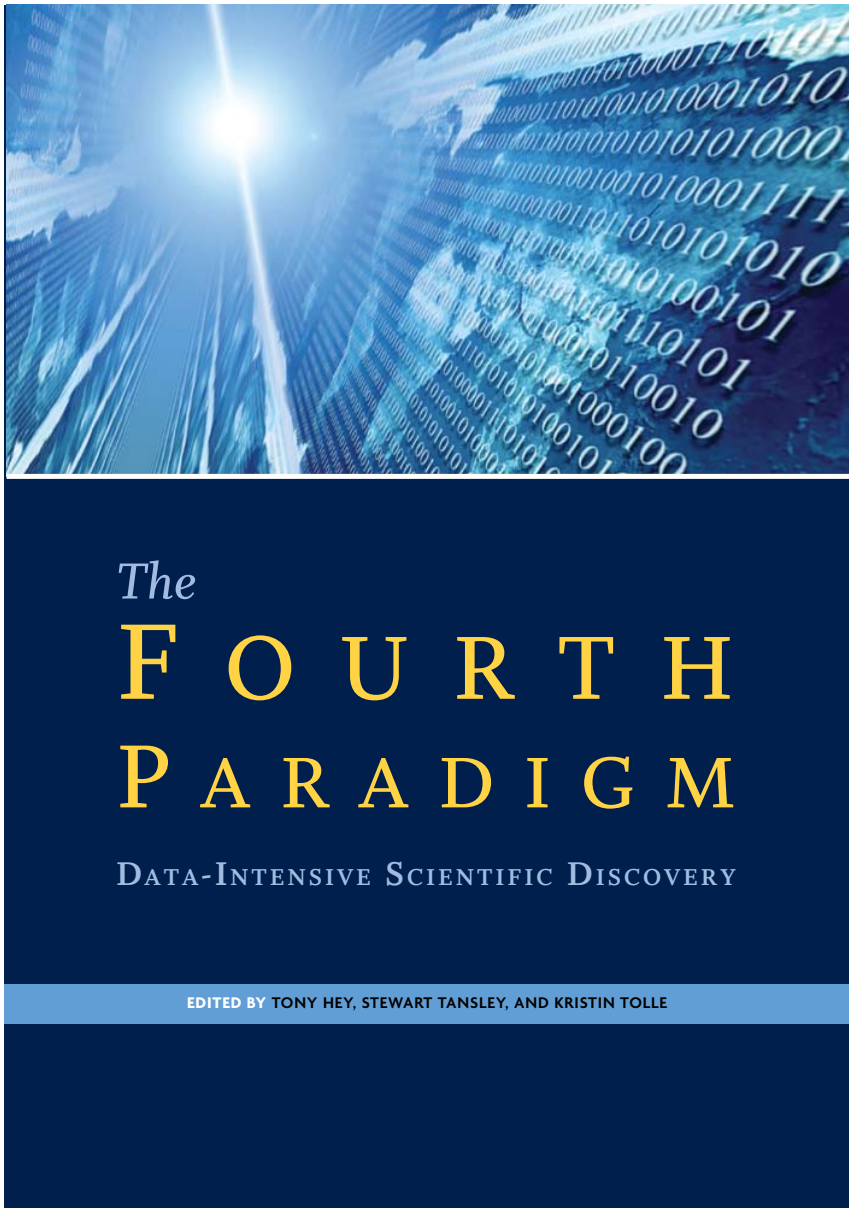
- Capability models

Community Capability Model Framework (CCMF)

- Collaboration
- Skills and training
- Openness
- Technical infrastructure
- Common practices
- Economic and business models
- Legal and ethical issues
- Academic culture

Credits

Background

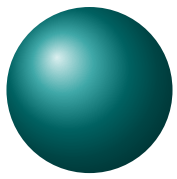


Microsoft® Research

"This book presents the first broad look at the rapidly emerging field of data-intensive science, with the goal of influencing the worldwide scientific and computing research communities and inspiring the next generation of scientists."

http://research.microsoft.com/en-us/collaboration/fourthparadigm/

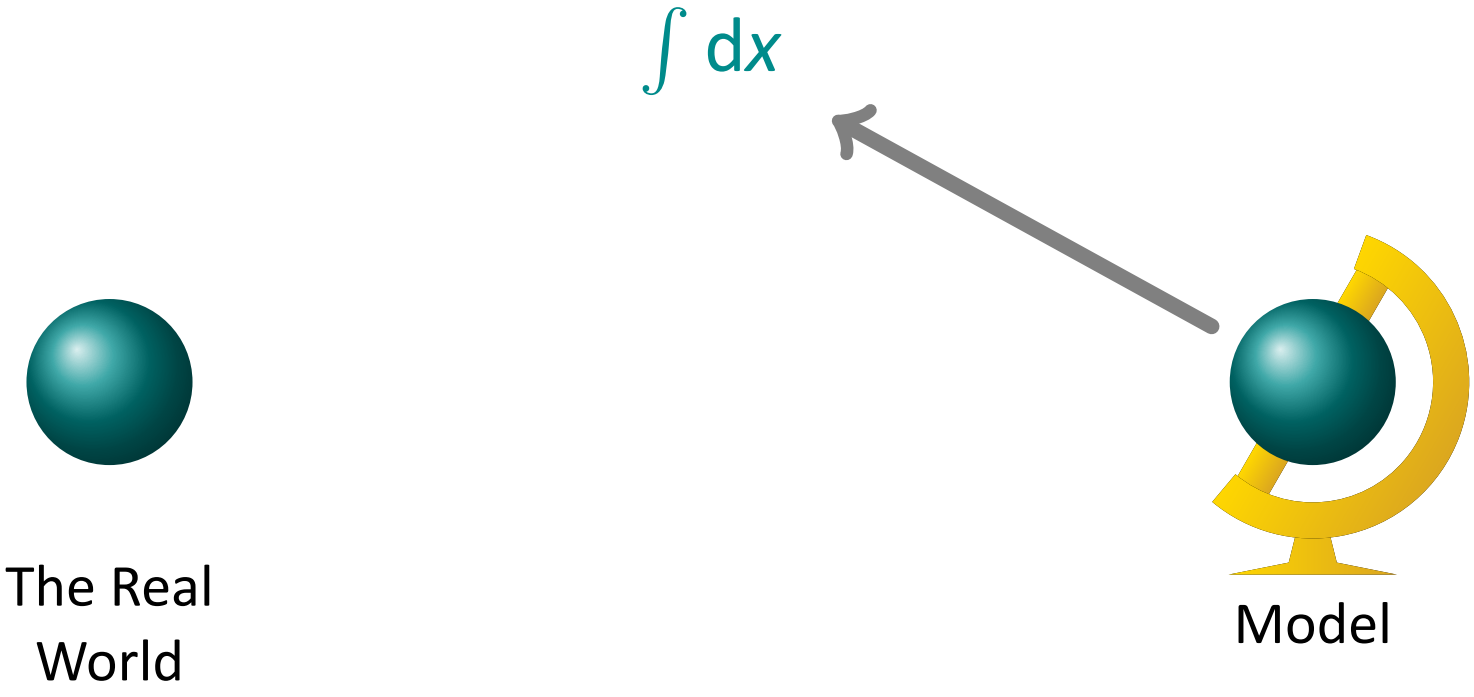
Four paradigms



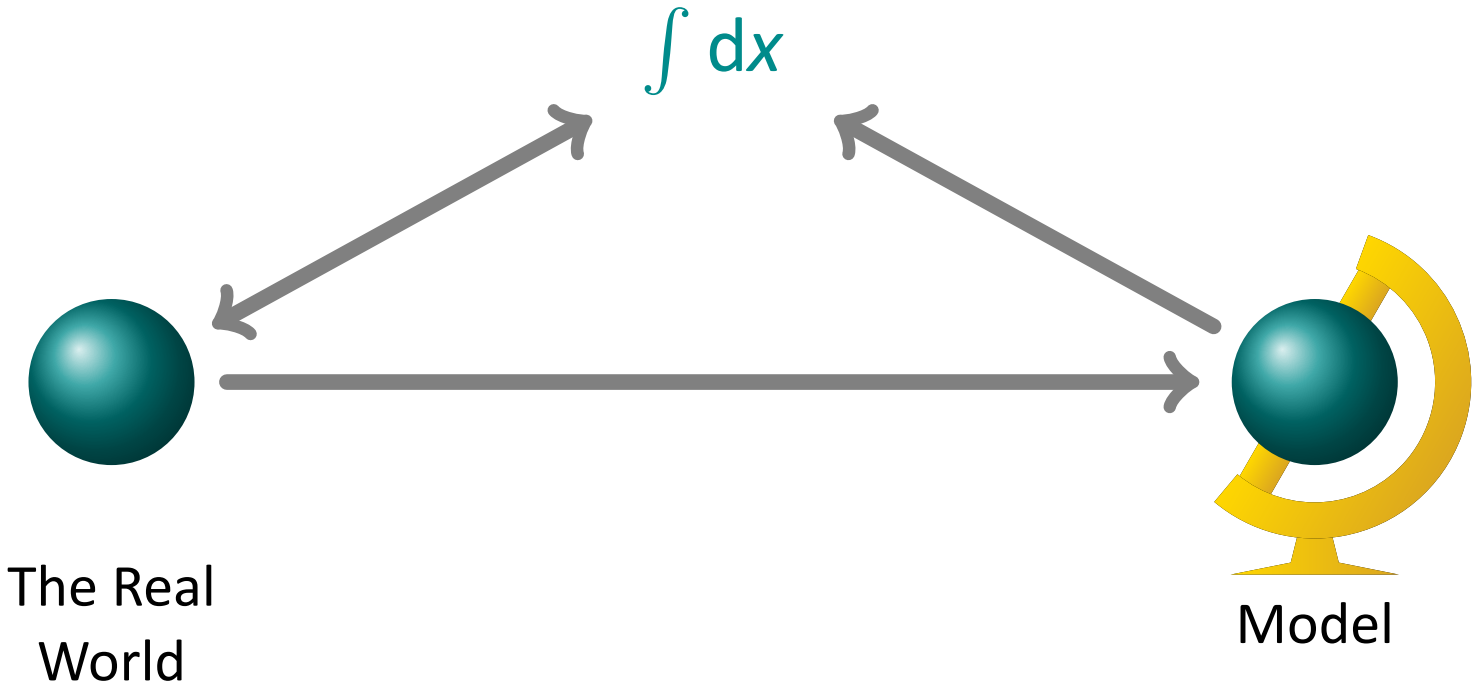
The Real World

Four paradigms

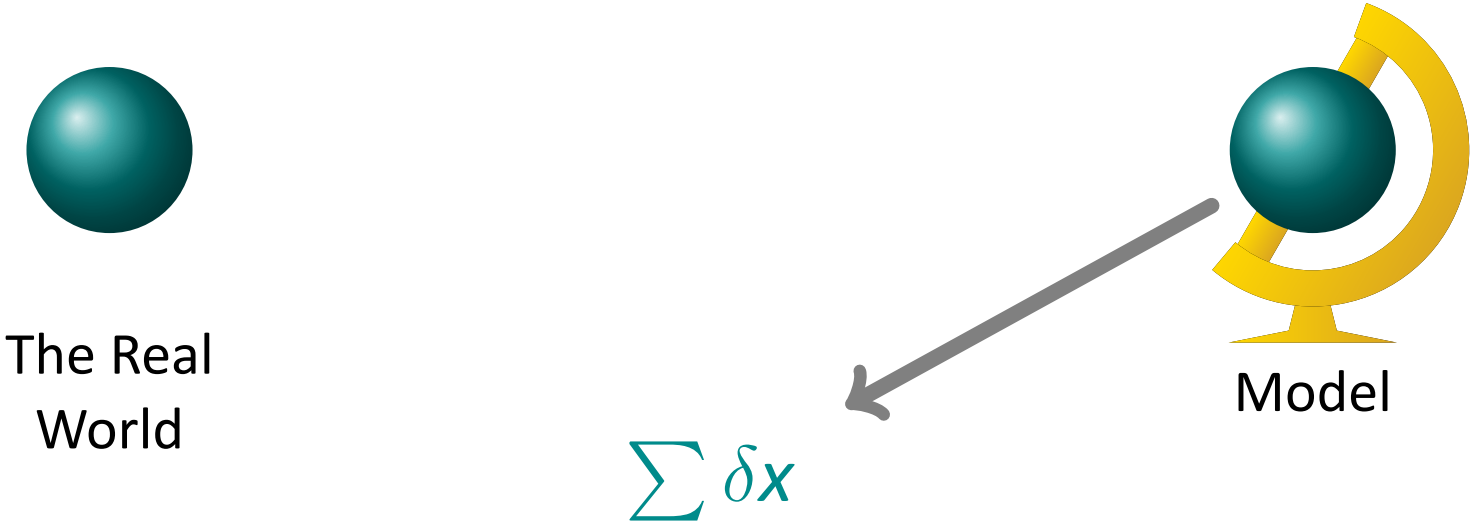
Four paradigms



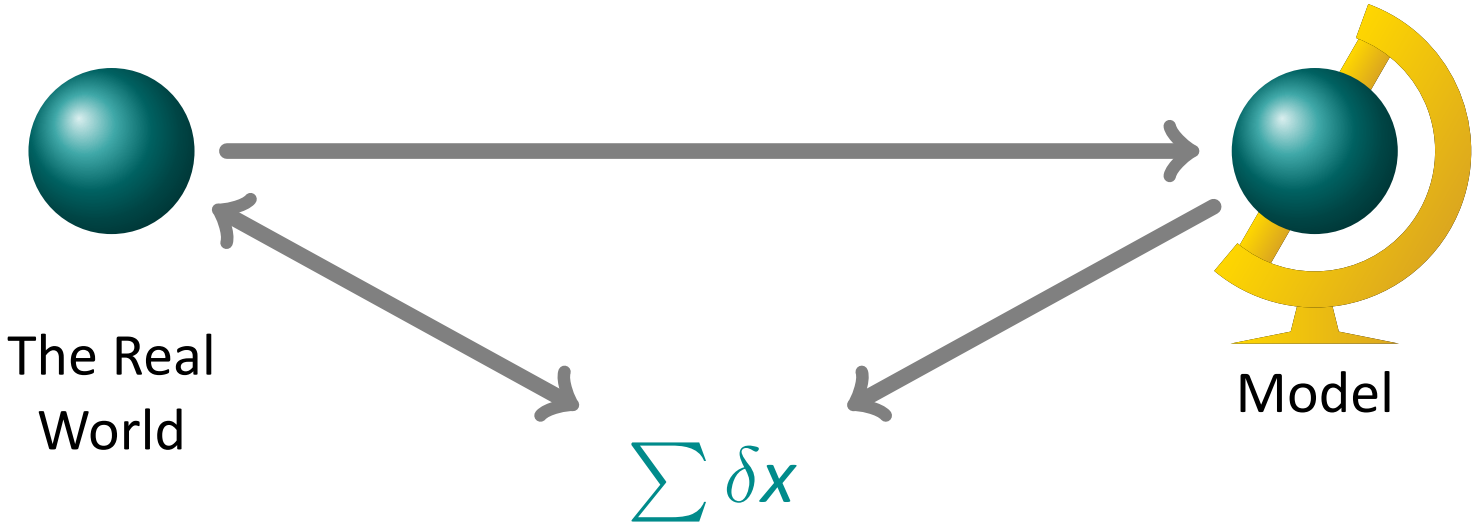
Four paradigms



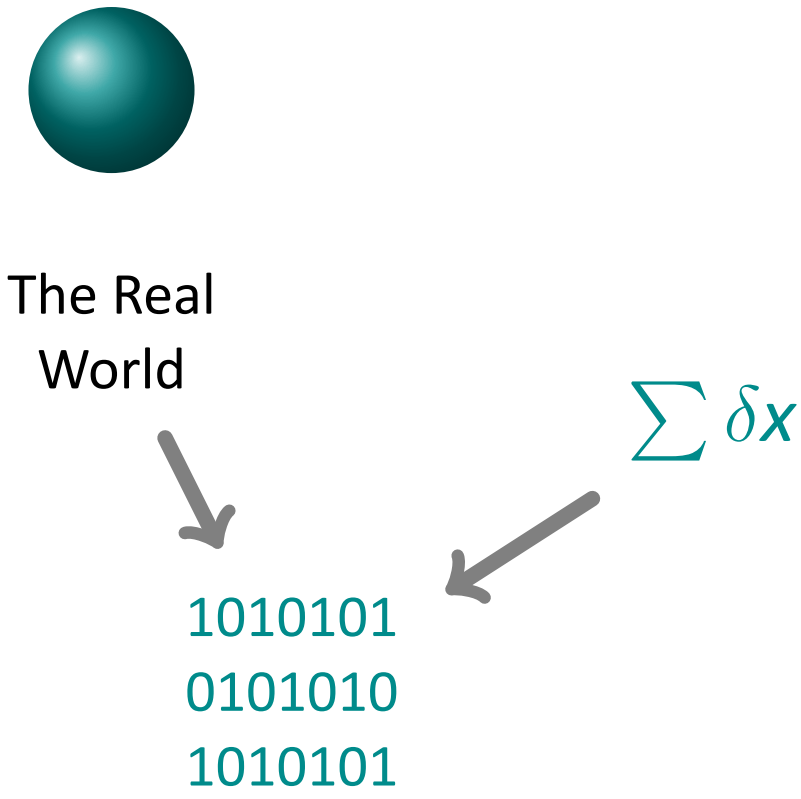
Four paradigms



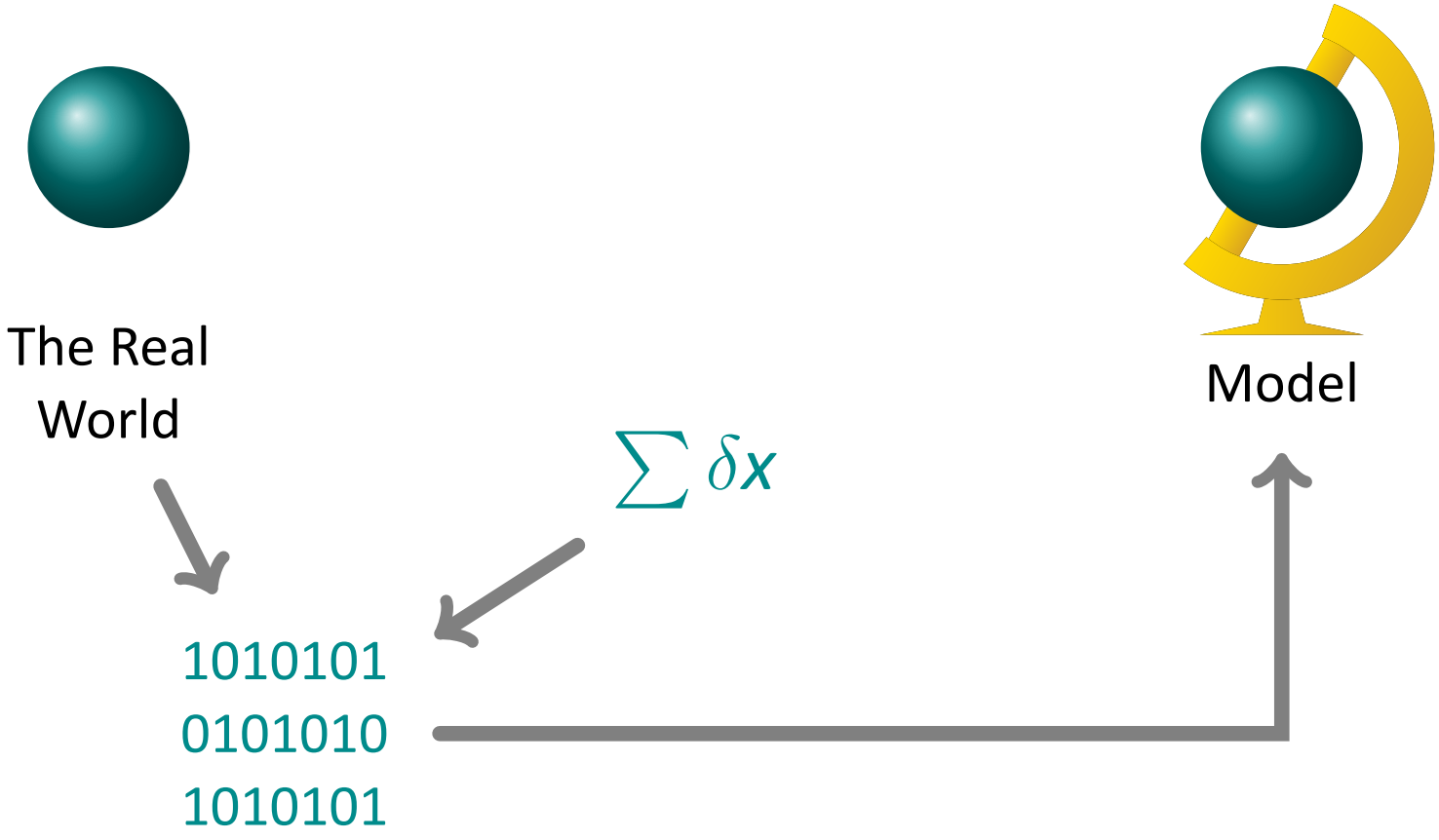
Four paradigms



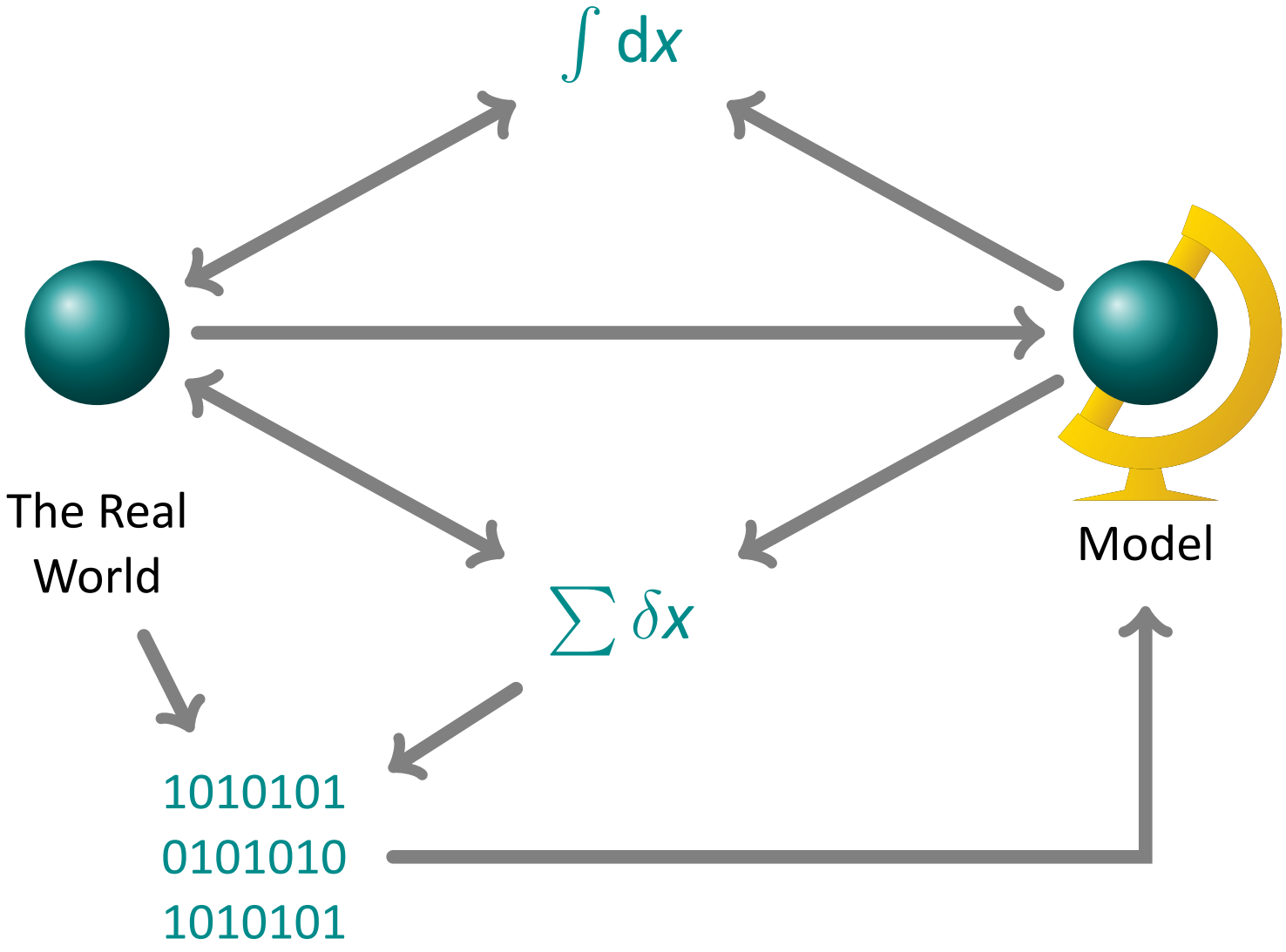
Four paradigms



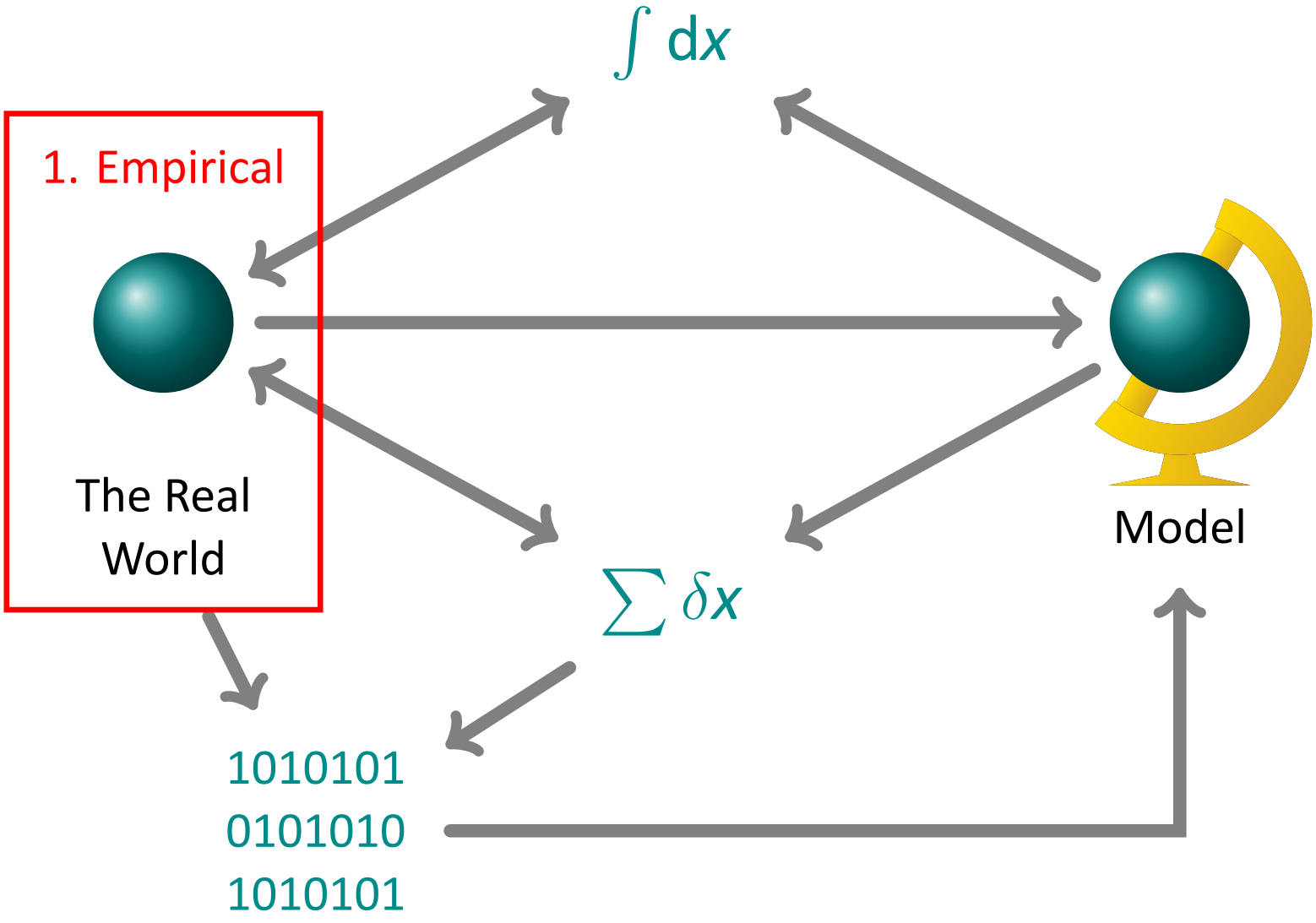
Four paradigms



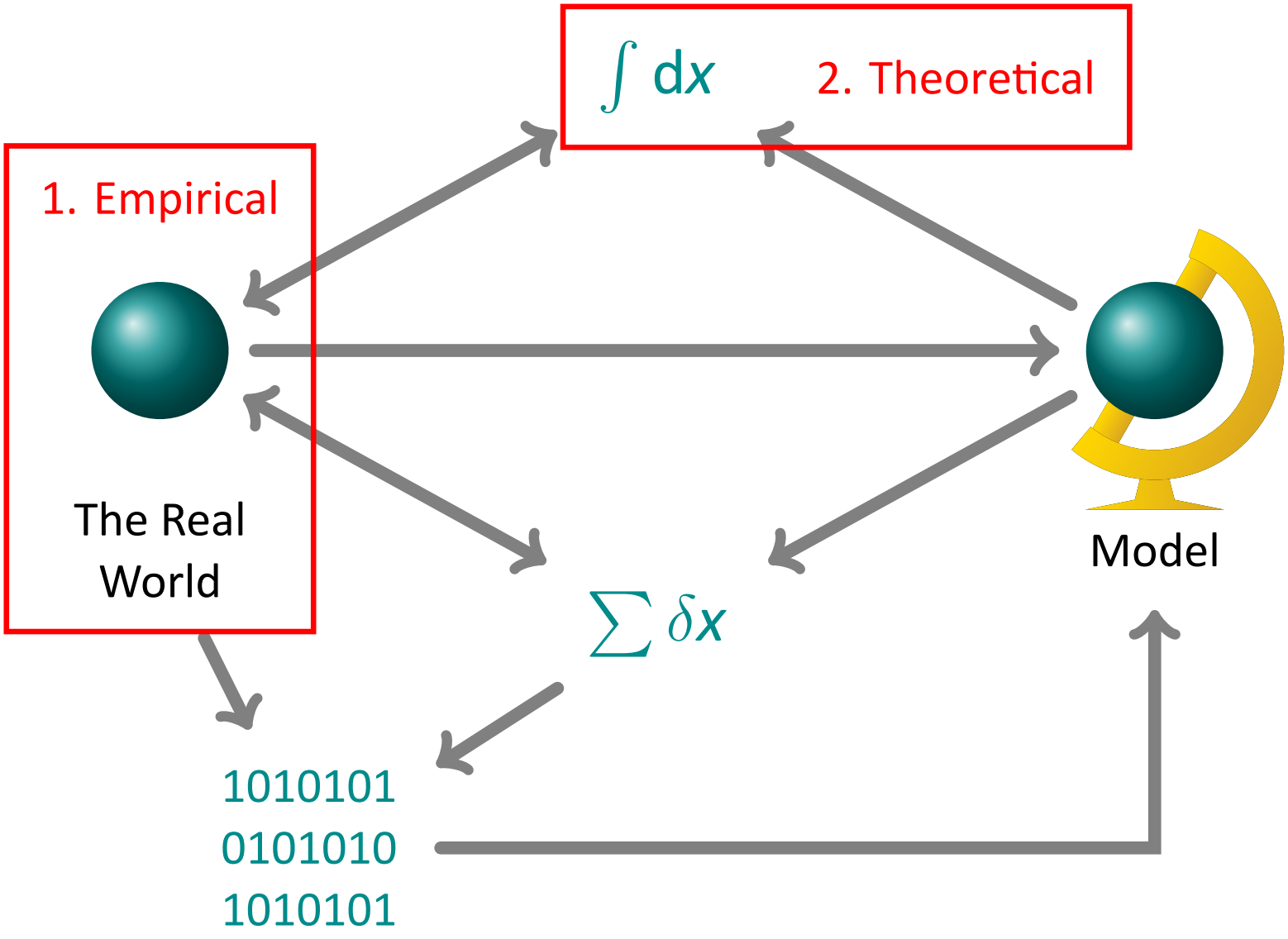
Four paradigms



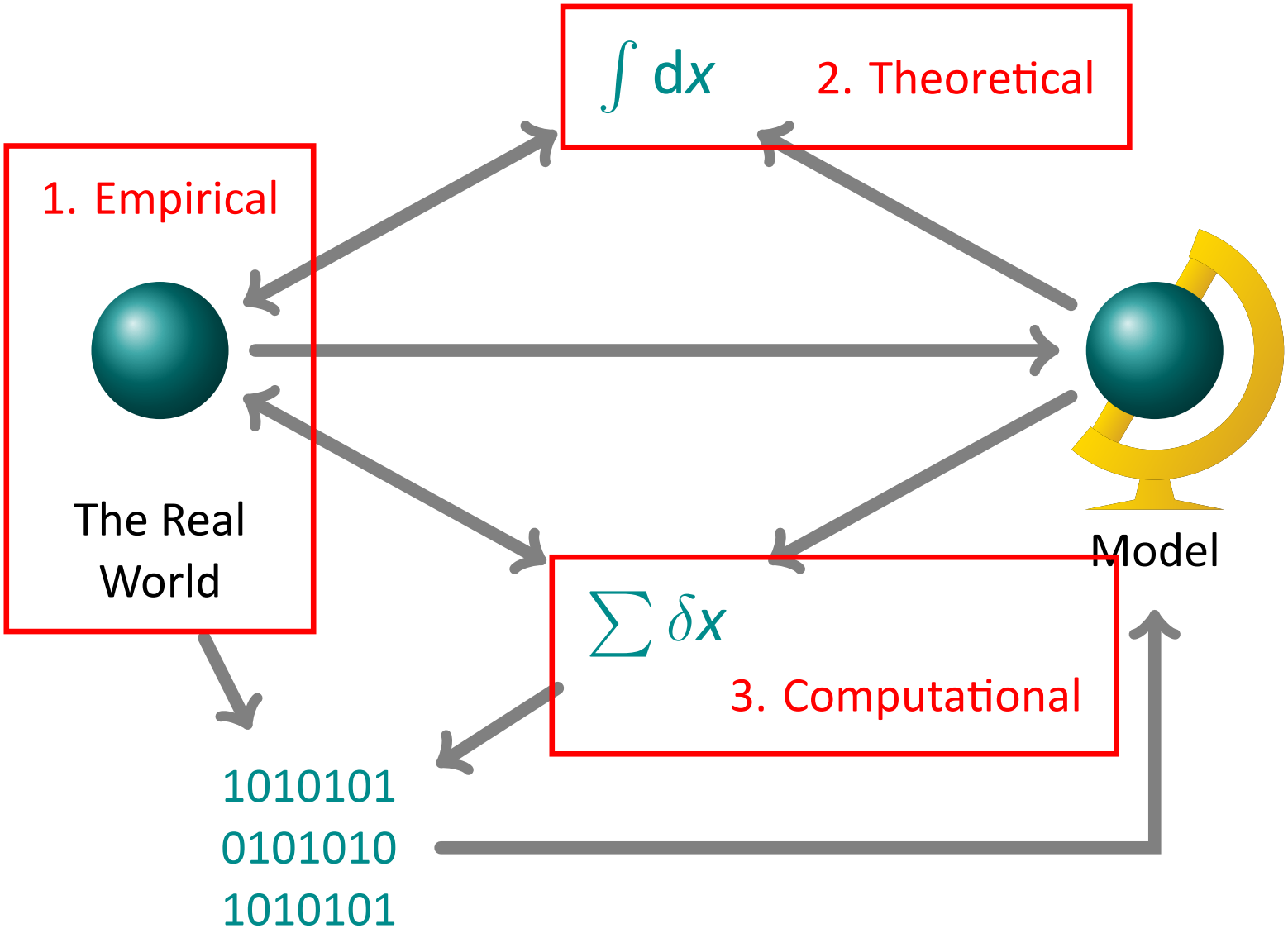
Four paradigms



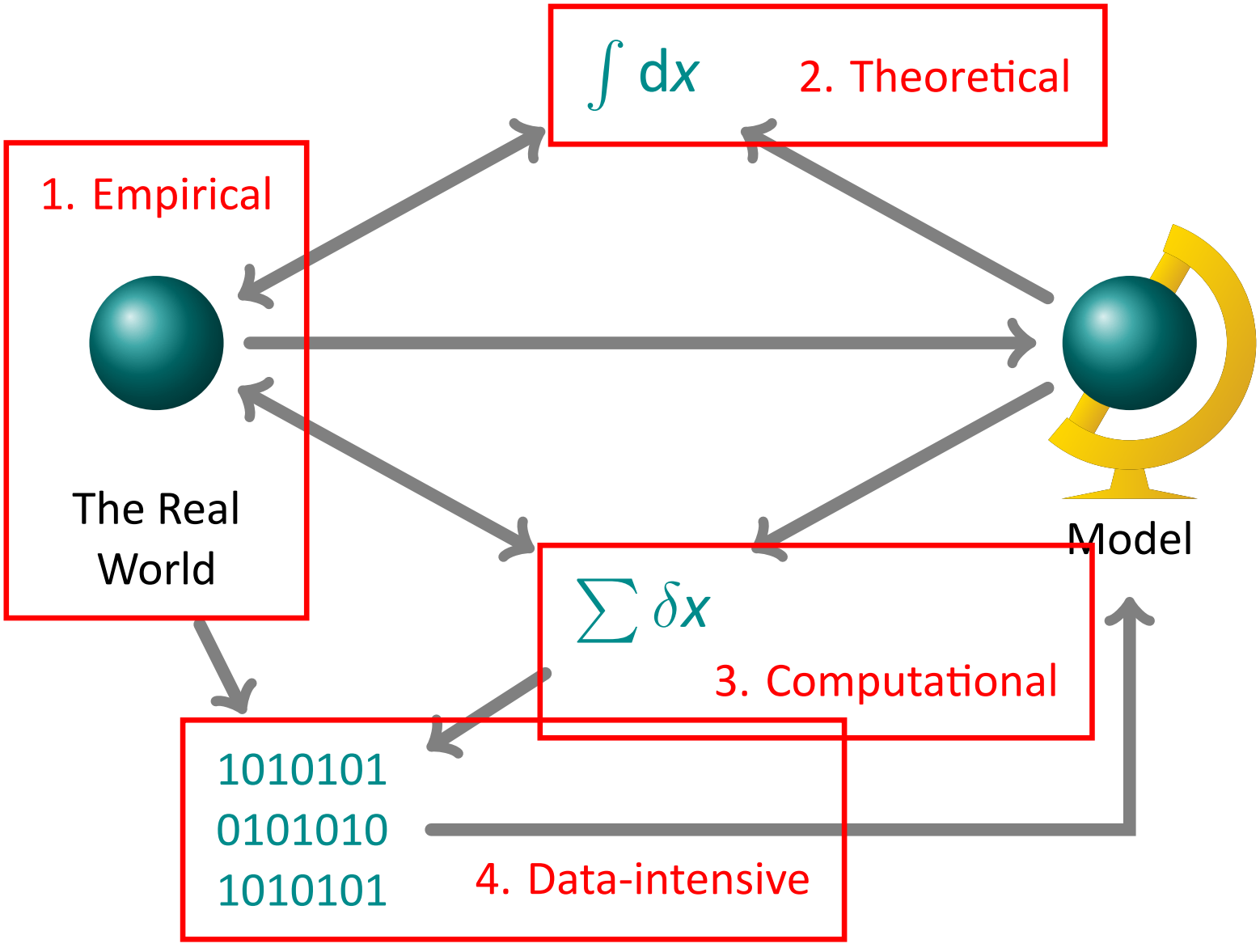
Four paradigms



Four paradigms



Four paradigms



Project objectives

1. Understand diversity in data-driven research *(consult)*
2. Identify and deconstruct factors contributing to community capability *(scope)*
3. Explore modes, metrics and dimensions for the capability factors *(describe)*
4. Develop a Community Capability Model Framework *(model, visualise)*
5. Produce domain mini case studies and business usage cases *(validate)*

Liz Lyon *(alt.)*

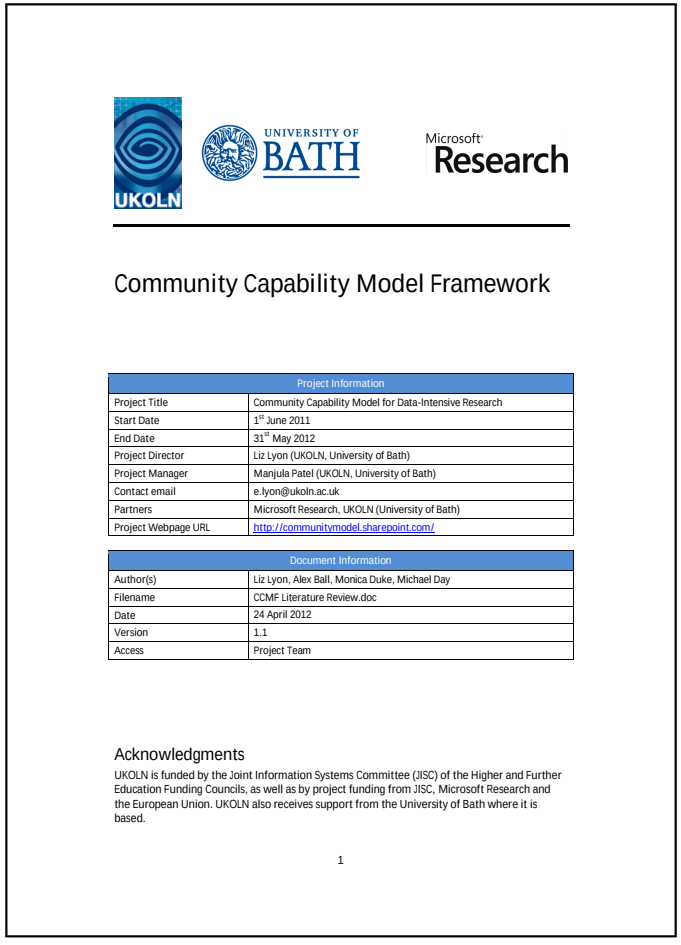
The White Paper

Purpose

The Community Capability Model Framework (CCMF) will provide support for:

- Intelligence-gathering
- Decision-making
- Planning
- Investment
- Building capacity
- Building capability
- Knowledge transfer

Liz Lyon



UKOLN | University of Bath | Microsoft Research

Community Capability Model Framework

Project Information	
Project Title	Community Capability Model for Data-Intensive Research
Start Date	1st June 2011
End Date	31st May 2012
Project Director	Liz Lyon (UKOLN, University of Bath)
Project Manager	Manjula Patel (UKOLN, University of Bath)
Contact email	e.lyon@ukoln.ac.uk
Partners	Microsoft Research, UKOLN (University of Bath)
Project Webpage URL	http://communitymodel.sharepoint.com/

Document Information	
Author(s)	Liz Lyon, Alex Ball, Monica Duke, Michael Day
Filename	CCMF Literature Review.doc
Date	24 April 2012
Version	1.1
Access	Project Team

Acknowledgments

UKOLN is funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from JISC, Microsoft Research and the European Union. UKOLN also receives support from the University of Bath where it is based.

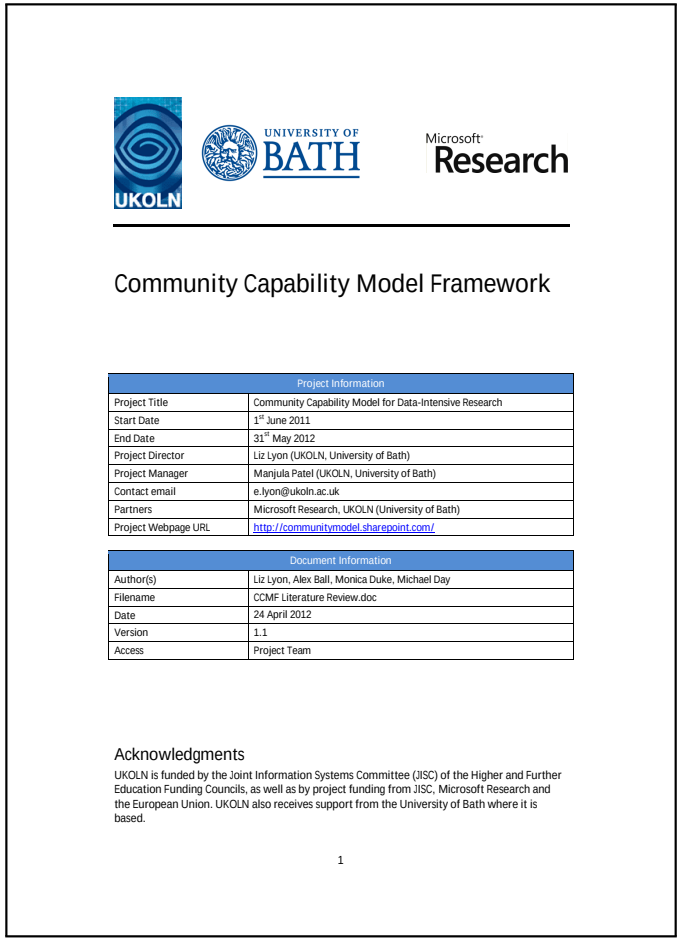
1

Purpose

The Community Capability Model Framework (CCMF) will provide support for:

- Intelligence-gathering
- Decision-making
- Planning
- Investment
- Building capacity
- Building capability
- Knowledge transfer

Liz Lyon



UKOLN | University of Bath | Microsoft Research

Community Capability Model Framework

Project Information	
Project Title	Community Capability Model for Data-Intensive Research
Start Date	1st June 2011
End Date	31st May 2012
Project Director	Liz Lyon (UKOLN, University of Bath)
Project Manager	Manjula Patel (UKOLN, University of Bath)
Contact email	e.lyon@ukoln.ac.uk
Partners	Microsoft Research, UKOLN (University of Bath)
Project Webpage URL	http://communitymodel.sharepoint.com/

Document Information	
Author(s)	Liz Lyon, Alex Ball, Monica Duke, Michael Day
Filename	CCMF Literature Review.doc
Date	24 April 2012
Version	1.1
Access	Project Team

Acknowledgments

UKOLN is funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from JISC, Microsoft Research and the European Union. UKOLN also receives support from the University of Bath where it is based.

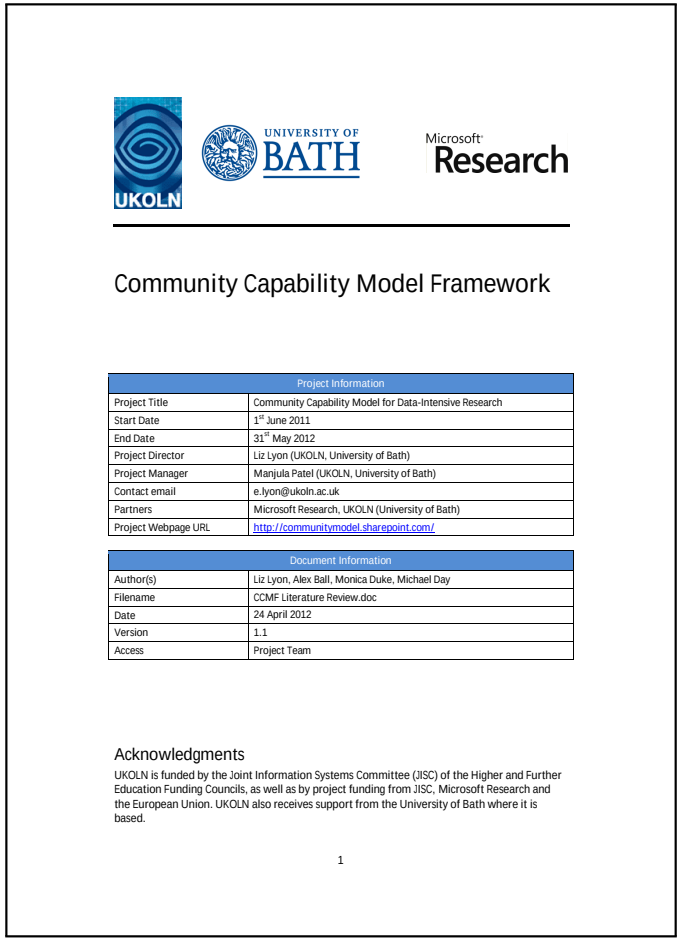
1

Purpose

The Community Capability Model Framework (CCMF) will provide support for:

- Intelligence-gathering
- Decision-making
- Planning
- Investment
- Building capacity
- Building capability
- Knowledge transfer

Liz Lyon



UKOLN | University of Bath | Microsoft Research

Community Capability Model Framework

Project Information	
Project Title	Community Capability Model for Data-Intensive Research
Start Date	1st June 2011
End Date	31st May 2012
Project Director	Liz Lyon (UKOLN, University of Bath)
Project Manager	Manjula Patel (UKOLN, University of Bath)
Contact email	e.lyon@ukoln.ac.uk
Partners	Microsoft Research, UKOLN (University of Bath)
Project Webpage URL	http://communitymodel.sharepoint.com/

Document Information	
Author(s)	Liz Lyon, Alex Ball, Monica Duke, Michael Day
Filename	CCMF Literature Review.doc
Date	24 April 2012
Version	1.1
Access	Project Team

Acknowledgments

UKOLN is funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from JISC, Microsoft Research and the European Union. UKOLN also receives support from the University of Bath where it is based.

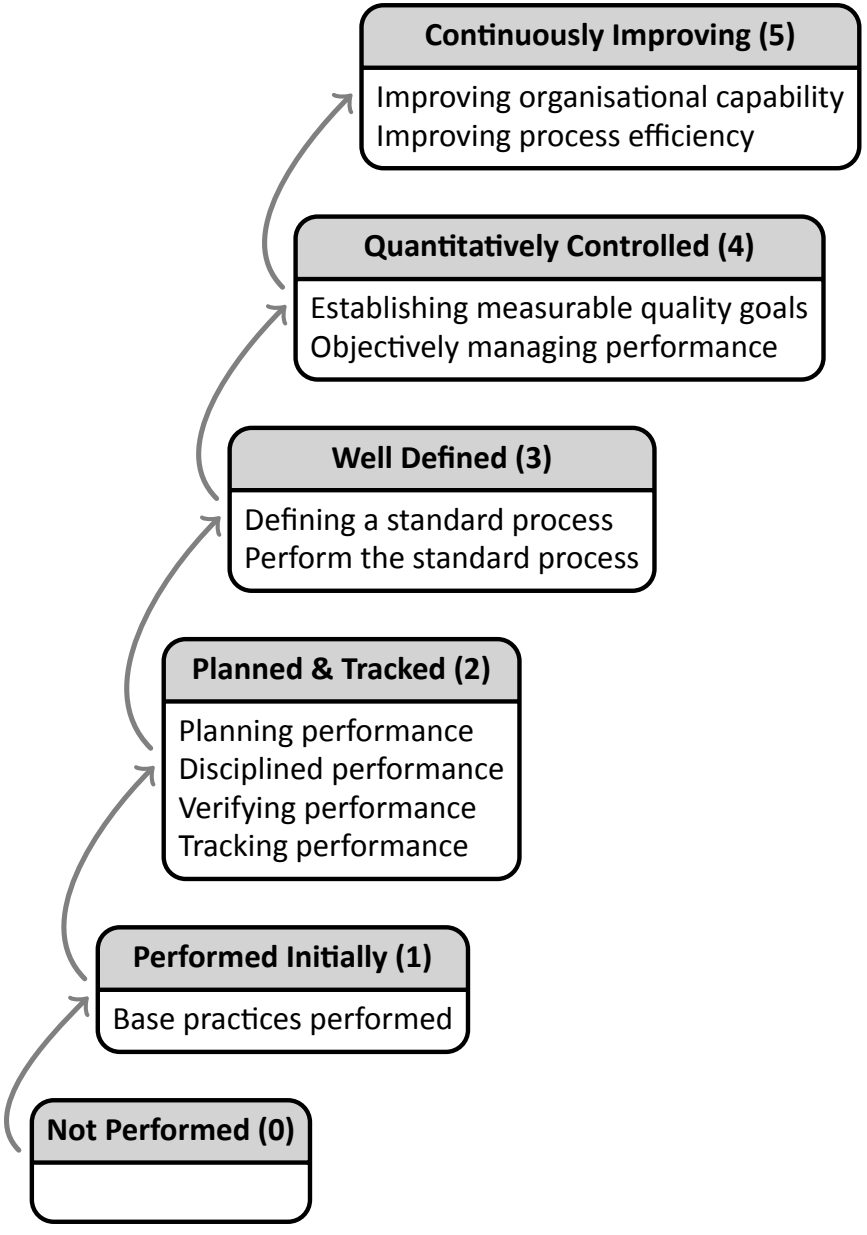
1

Community Maturity Model

	Stage 1 Hierarchy	Stage 2 Emergent Community	Stage 3 Community	Stage 4 Network
Strategy	Familiarize & Listen	Participate	Build	Integrate
Leadership	Command & Control	Consensus	Collaborative	Distributed
Culture	Reactive	Contributive	Emergent	Activist
Community Management	None	Informal	Defined roles & processes	Integrated roles & processes
Content & Programming	Formal & Structured	Some user generated content	Community created content	Integrated formal & user generated
Policies & Governance	No Guidelines	Restrictive	Flexible	Inclusive
Tools	Consumer tools used by individuals	Consumer & self-service tools	Mix of consumer & enterprise tools	'Social' functionality is integrated throughout
Metrics & Measurement	Anecdotal	Activity Tracking	Activities & Content	Behaviors & Outcomes

http://community-roundtable.com/2009/06/the-community-maturity-model/

Systems Engineering Capability Maturity Model



- The enterprise is divided into *process areas* (e.g. Ensure Quality, Manage Risk).
- Achieving a *capability level* within a process area means implementing a certain set of practices.
- These practices are grouped into *common features* (see figure).
- At Level 1, each process area has its own set of *base practices*.
- At Levels 2--5, all process areas share sets of *generic practices*.

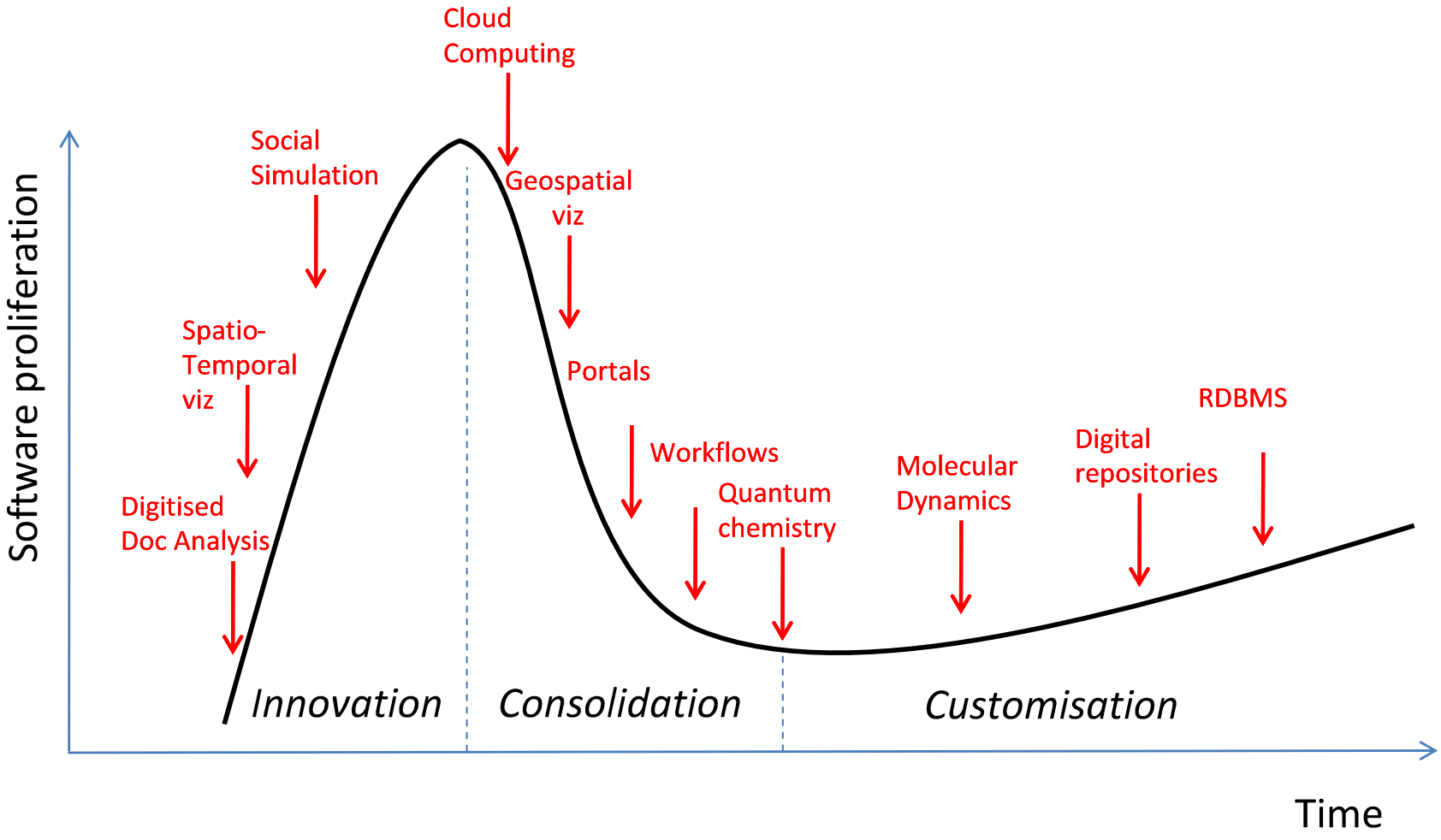
`http://www.sei.cmu.edu/reports/95mm003.pdf`

Three-Legged Stool

Organisation	Technology	Resources
1. Data Ownership and Management	1. Technological Infrastructure	1. Data Management Costs and Sustainability
2. Data Policies and Procedures	2. Appropriate Technologies	2. Business Planning
3. Data Policy Review	3. Ensuring Availability	3. Technological Resources Allocation
4. Sharing of Research Data/Access to Research Data	4. Managing data integrity	4. Risk Management
5. Preservation and Continuity of Research	5. Obsolescence	5. Transparency of Resource Allocation
6. Internal Audit of Research Activities	6. Managing technological change	6. Sustainability of Funding for Data Management and Preservation
7. Monitoring and Feedback of Publication	7. Security Provisions	7. Data Management Skills
8. Metadata Management	8. Security Processes	8. Number of Staff for Data Management
9. Legal Compliance	9. Metadata tools	9. Staff Development Opportunities
10. Intellectual Property Rights and Rights Management	10. Institutional Repository	
11. Disaster Planning and Continuity of Research		

http://cardio.dcc.ac.uk/

Software Maturity Curve



https://ocs.arcs.org.au/index.php/eraust/2010/paper/view/141

Community Capability Model Framework (CCMF)

UNIVERSITY OF BATH

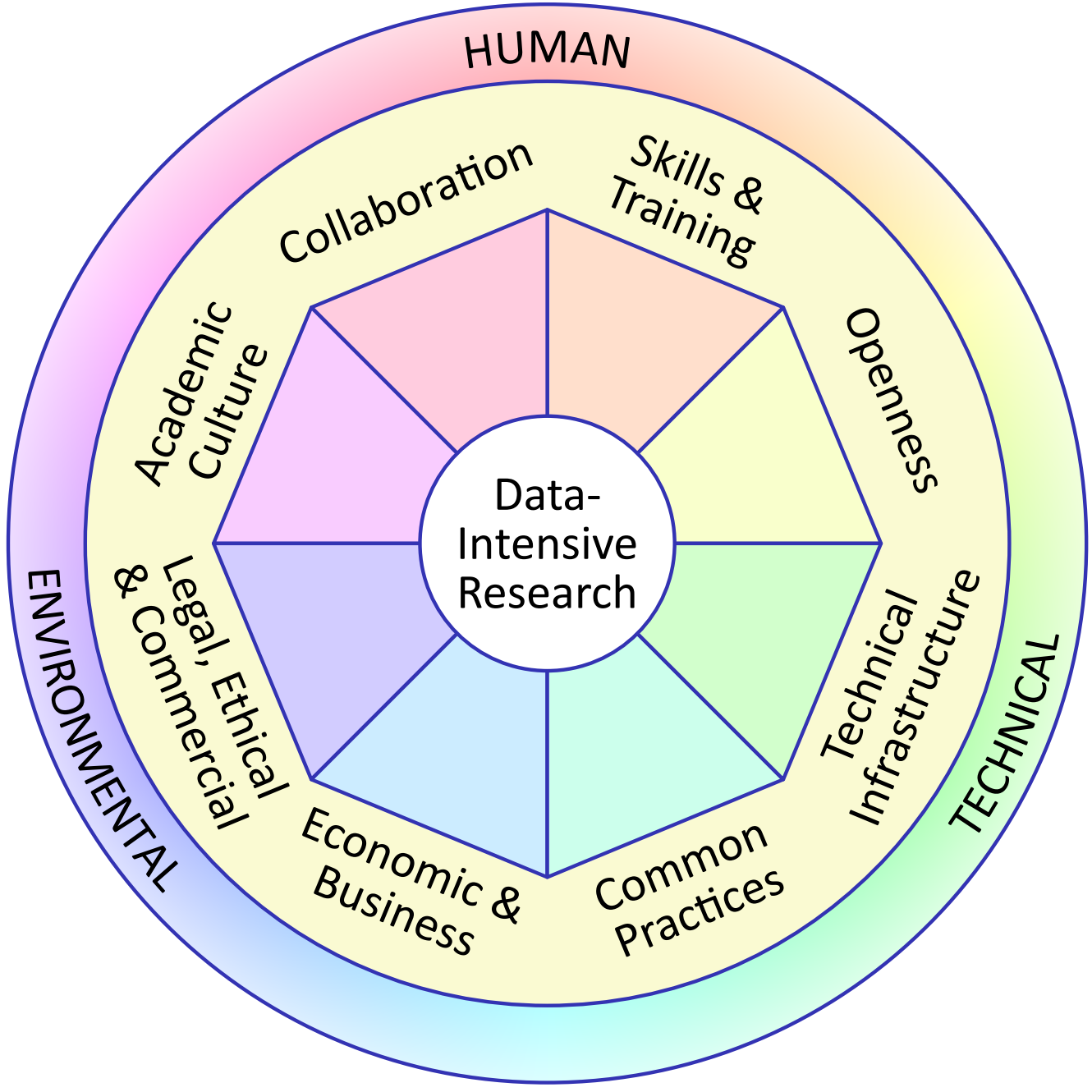
Consultation

Case studies

- ESRC
- P.I.s and research leaders from eResearch South Consortium
- University of Bath (Pro-VC for Research, Computing Services, Research Office,…)

Workshops

- York
- Harvard
- Bristol
- Stockholm
- Melbourne


HUMAN
Collaboration
Skills & Training
Openness
Technical Infrastructure
Common Practices
Economic & Business
Legal, Ethical & Commercial
Academic Culture
Data-Intensive Research
ENVIRONMENTAL
TECHNICAL

Collaboration

Collaboration within the discipline/sector

Lone researchers.	Departmental research groups.	Collaboration across research groups within or between organisations.	Discipline organised at a national level.	International collaboration and consortia.

Collaboration and interaction across disciplines

No collaboration with other disciplines.	Individual researchers occasionally collaborate outside their discipline.	Disciplines collaborate through joint conferences or publications.	Bilateral collaborations.	Formal collaboration between research groups from several different disciplines.

Collaboration (cont.)

Collaboration and interaction across sectors

None.	Attempts have been made but are not considered successful.	Despite successful examples working with other sectors is not the norm – some barriers are perceived.	A discipline or group has gained experience of working closely with one or two sectors.	Work successfully with several other sectors on different problems.

Collaboration with the public

No collaboration with the public.	The public's involvement is limited to acting as subjects of study, user testing, etc.	Contact with the public is only through occasional appearance in the media e.g. news bulletins, TV programmes.	Mainly informational, sometimes participative, targeted media programmes are organised to engage the public e.g. science fairs.	Dedicated programmes involving the public in research; crowd sourcing/ citizen science.

Skills and training

Skill sets

Tools and technologies (cloud computing, visualisations, statistical analysis, simulations, modelling).	Data description and identification (metadata, vocabularies, citation).	Collaboration and communication (engaging with other researchers, the public, the media).	Policy and planning (data management, business models).

Pervasion of training

No training avail-able.	Training pro-grammes in devel-opment.	Training available but not embedded within u/g and p/g degree programmes. Patchy uptake. Little or no on-job coaching or mentoring on data management.	Training embedded within u/g and p/g degree programmes and available for researchers. Mentors usually provided on request.	Dedicated training, fully embedded in all u/g and p/g degree programmes, accredited with professional qualifications, and an established part of continuing professional development.

Openness

Openness in the course of research

No sharing. No details released.	Selected details released, e.g. in a proposal or project plan.	Selected intermediate results are shared within a limited group.	Intermediate results are shared through traditional means, e.g. conference papers.	Sharing is done publicly on the web. Full details are disclosed.

Openness of published literature

No sharing of papers or metadata outside publication channels.	Authors share metadata for their publica-tions.	Authors share theses or other selected sections from the literature.	Authors provide copies of their publications on request or other negotiated means.	Publica-tions are made available on open access.

Openness (cont.)

Openness of data

No sharing. No details released.	The data are described in the literature but not made available.	Data are available on request, after embargo or with other conditions.	Efforts are made to make data discoverable and re-usable as well as available.	Data is available in re-usable form and freely available to all. Community curation of the data may be possible.

Openness of methodologies/workflows

No sharing. No details released.	Released within limited scope. Partial details released.	The details of the workflow are shared but not the underlying scripts; only partial stages of the workflow are shared.	Sharing publicly on the web. Non-standard scripts, tools and software released.

Reuse of existing data

Only own data used.	Data exchanged within limited scope.	Regularly combine data sets in specific established ways. Provenance tracked in ad hoc ways.	Multiple existing datasets often combined. Provenance tracked systematically.

Technical infrastructure

Computational tools and algorithms

None.	Tools exist but perform below requirements.	Tools have sufficient features to meet the needs of most users.	Tools have features few people use, expected to meet users' needs for the next few years.

Tool support for data capture and processing

No tool support for data capture.	Tools do not meet user requirements well or do not interoperate. Tools are custom and quality varies.	One or two good tools available. A few clear leaders.	Most tools that support data capture do it well and meet user requirements.	All tools support data capture well and interoperate. There is a good choice of tools for data processing.

Data storage

None.	Insufficient data storage available to meet user needs.	Although data storage capacity is sufficient, other requirements (e.g. security) are not met.	Dedicated storage facilities meet current requirements, but will be outgrown shortly.	Storage is available and is expected to meet future needs.

Technical infrastructure (cont.)

Support for curation and preservation

None.	Support is only available in specialised cases.	Insufficient tools and facilities exist to meet needs.	Dedicated tools are available and are widely used.	Common infrastructure is well funded and well used.

Data discovery and access

None.	Discovery services very discipline-specific; require specialised knowledge or rights.	Discovery opened to all but siloed (not interopeable).	Data discoverable and accessible to all, good integrated services.

Integration and collaboration platforms

None.	Platforms exist but perform below requirements.	Platforms have sufficient features to meet the needs of most users.	Platforms have features few people use, expected to meet users' needs for the next few years.

Technical infrastructure (cont.)

Visualisations and representations

None.	Tools exist but perform below requirements.	Tools have sufficient features to meet the needs of most users.	Tools have features few people use, expected to meet users' needs for the next few years.

Platforms for citizen science

None.	Customised tools available, used by a small number of groups.	Very flexible tools available and well used.	Tools have been re-deployed to other disciplines.

Common practices

Data formats

No standard formats available: ad hoc formats prolifer-ate.	Standard formats are in develop-ment but not yet in use.	Some standard formats available but not widely adopted or community begins to converge on small number of formats.	Standard formats are widely adopted for some but not all types of data.	Standard formats are universally adopted for all types of data. Faithful conversions are possible between 'rival' standards.

Data collection methods

Methods are not usually shared.	Methods are shared but not widely reused.	Agreed methods are in develop-ment.	Although some methods are agreed there are gaps in the methods covered or room for improvement in the quality.	Methods are well known, well documented and well used.

Common practices (cont.)

Processing workflows

Work-flows are not usually shared.	Work-flows are shared but not widely reused.	Agreed workflows are in development, or community begins to converge on a small number of workflows.	Agreed workflows are available with some gaps, or room for improvement in quality.	Several standard-ised workflows widely used.

Data packaging and transfer protocols

Pack-aging and transfer per-formed ad hoc.	Standard protocols are in de-velopment but not yet in use.	Some standard protocols available but not widely adopted or community begins to converge on small number of protocols.	Some standard protocols available with some gaps, or room for improvement in quality.	One or two standard-ised formats/pro-tocols widely used.

Common practices (cont.)

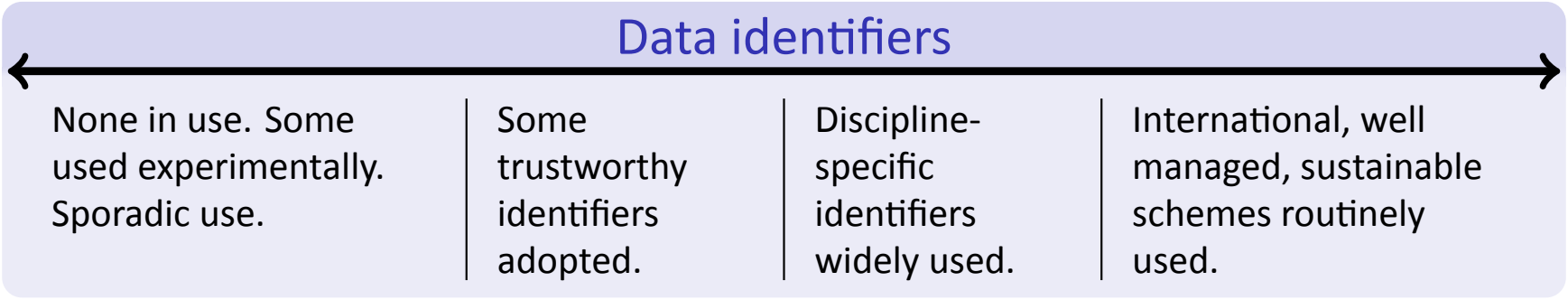
Data description

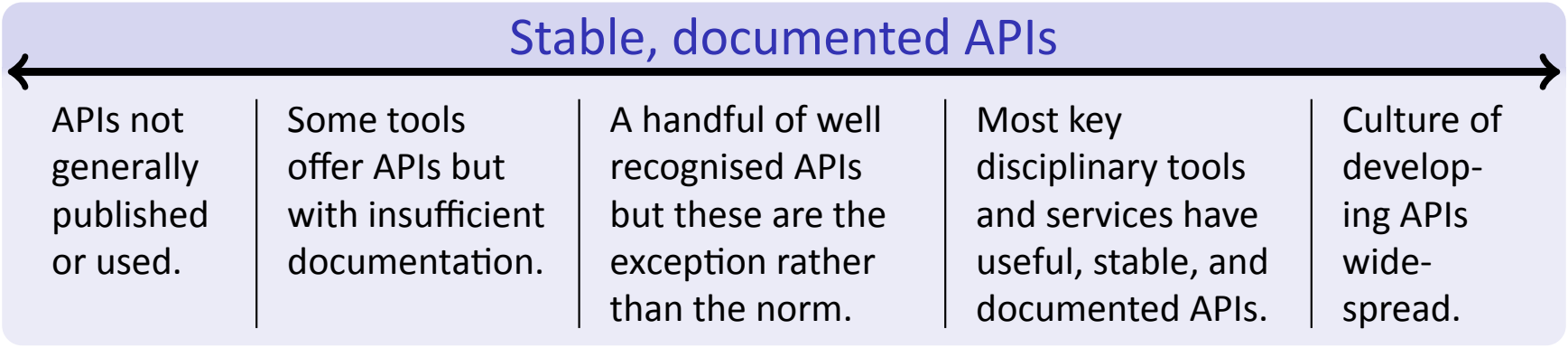
No standard metadata schemes exist.	Standard metadata schemes are in development but not yet in use.	Some metadata schemes are published and recognised, but with little uptake or known flaws.	Recognised metadata schemes agreed, with some gaps.	Mature, agreed and widely used metadata schemes exist.

Vocabularies, semantics, ontologies

No standard schemes are available.	Some schemes are published but they are experimental with limited uptake.	Standards are being actively developed; agreement and standardisation by the community is being pursued.	Some standard schemes are available, however gaps still exist.	Standard schemes are mature with good take-up by the community and widely applied.

Common practices (cont.)





Economic and business models

Sustainability of funding for research

Funding focused on short-term projects and quick returns.	Single-phase thematic investments on a 3-5 year timescale.	Multi-phase thematic investments in 5-10 year blocks which build a community.

Geographic scale of funding for research

Projects funded internally or through grants from regional agencies.	Projects funded by national funders.	Funding by international bodies and bi-lateral initiatives between national funders.

Size of funding for research

Small-scale projects (e.g. to exploit open innovation methodologies for bio-informatics tool development).	Mid-scale projects (e.g. digitisation and analysis of large textual corpora).	Major investment (e.g. in longitudinal data surveys).

Economic and business models (cont.)

Sustainability of funding for infrastructure

One-off investments with no commitment to sustainment.	Infrastructure projects allowed slow transition to self-financing model.	Sustained multi-decade investments in data centres and services.

Geographic scale of funding for infrastructure

Investments by a single funding body at regional or national level.	Collaborative development at the national level by multiple funders.	Collaborative development between international funders.

Size of funding for infrastructure

Small-scale tool development.	Co-ordinated investments in distributed systems.	Large central investments in network infrastructure or tools.

Economic and business models (cont.)

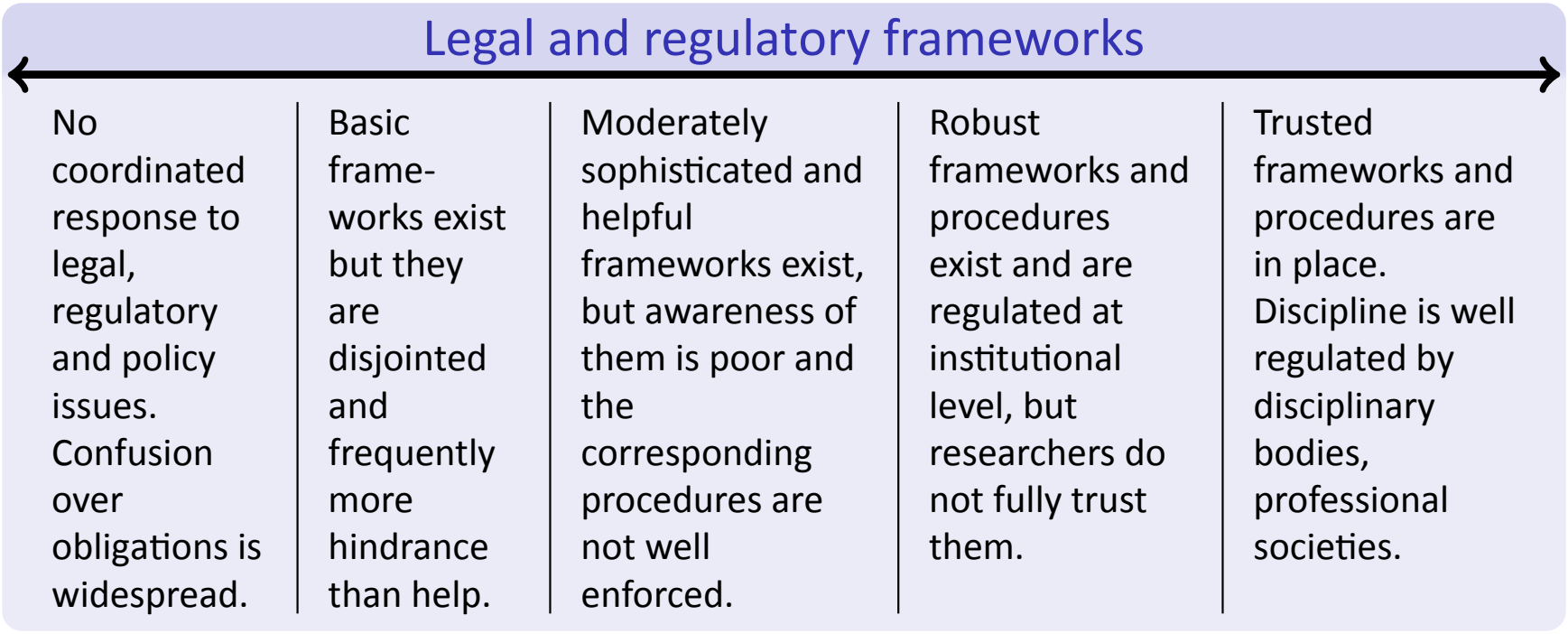
Public–private partnerships

None.	Informal collaboration with industry but no funding involved.	Corporate/SME are non-funded partners in proposals with academia.	Research is co-funded by industry and other sources.	Established formal co-investment partnerships running long-term multi-phase projects.

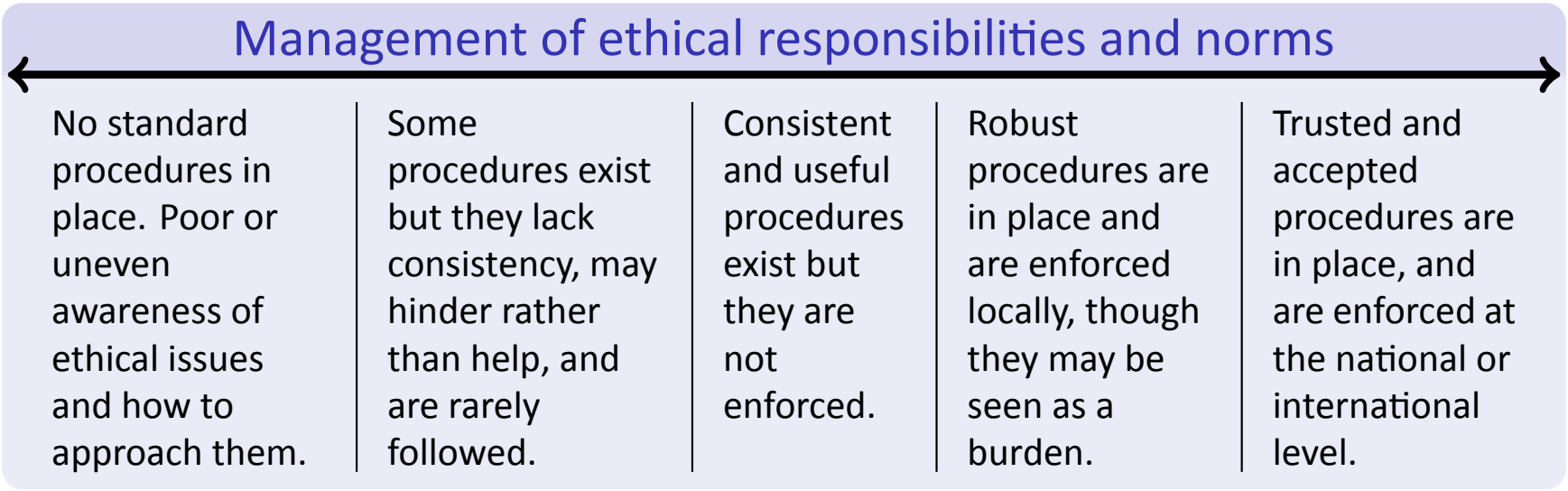
Productivity and return on investment

Long lead times between project start and submission of outputs (e.g. 6 years), and between acceptance and publication of papers (e.g. 2 years). Funders expect projects to publish a small number of papers each with high long-term impact.	Mid-range lead times between project start and submission of outputs (e.g. 3 years), and between acceptance and publication of papers (e.g. 1 year). Funders expect projects to publish a moderate number of papers in high impact journals.	Short lead times between project start and submission of outputs (e.g. 18 months), and between acceptance and publication of papers (e.g. 3 months). Funders expect projects to publish a large number of both high quality papers and progress reports.

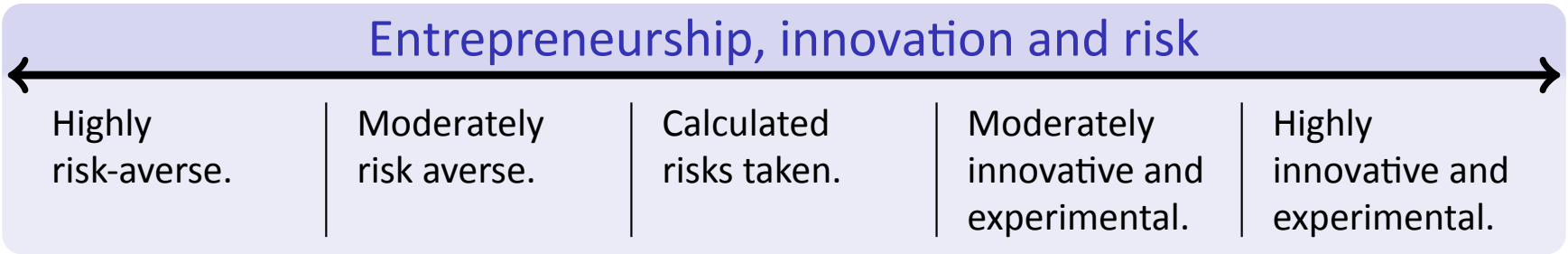
Legal and ethical issues

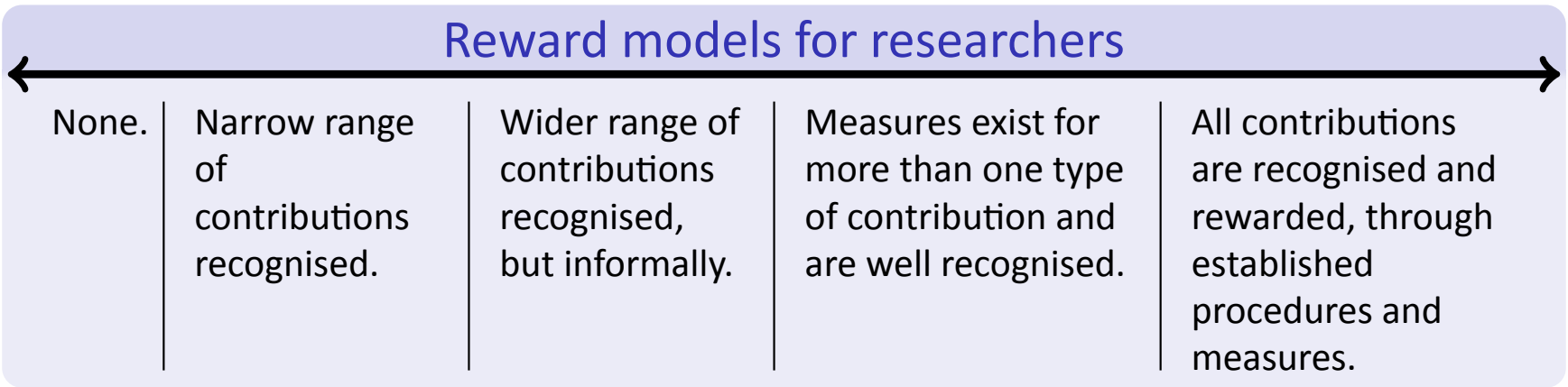


Legal and ethical issues (cont.)

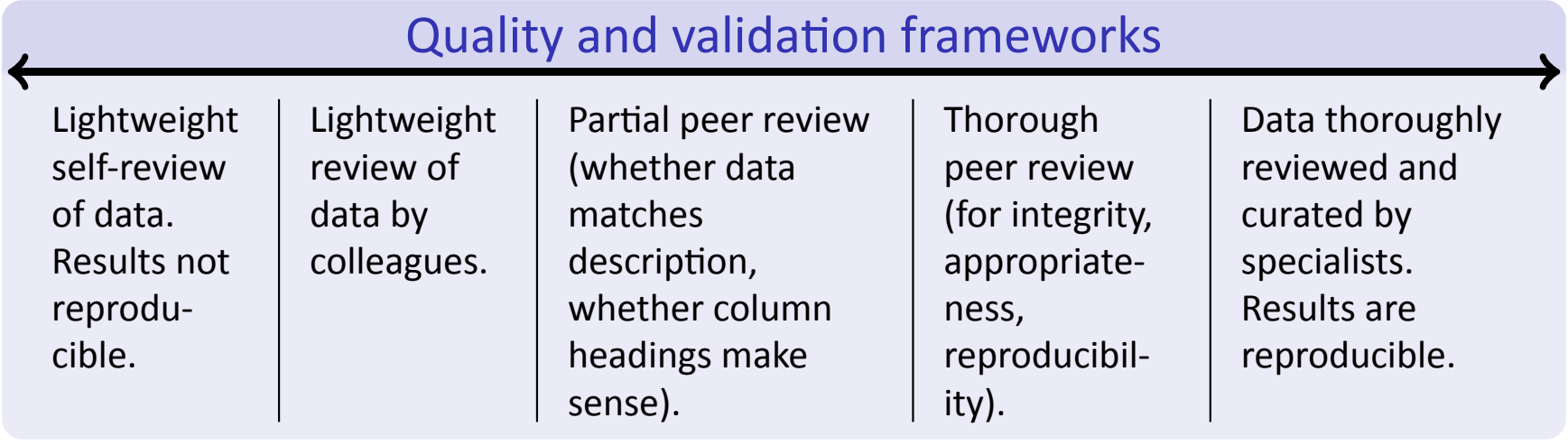


Academic culture





Academic culture (cont.)



DEMO

Credits

Project Team

Liz Lyon
Alex Ball
Michael Day
Monica Duke
Manjula Patel
Michelle Smith

Microsoft Research

Kenji Takeda
Alex Wade

Website: `http://communitymodel.sharepoint.com/`

Acknowledgements

The CCMDIR Project is funded by Microsoft Research.

UKOLN is funded by JISC, as well as by project funding from JISC, Microsoft Research and the European Union. UKOLN also receives support from the University of Bath where it is based.